AD NO. ___________________________
DTC PROJECT NO. 8-CO-160-UXO-021
REPORT NO. ATC 10071

STANDARDIZED

UXO TECHNOLOGY DEMONSTRATION SITE

SCORING RECORD NO. 923

SITE LOCATION:
U.S. ARMY ABERDEEN PROVING GROUND

DEMONSTRATOR:
NAVAL RESEARCH LABORATORY (NRL)
CODE 6110
WASHINGTON, DC 20375

TECHNOLOGY TYPE/PLATFORM:
EM61/PUSHCART

AREAS COVERED:
INDIRECT FIRE

PREPARED BY:
U.S. ARMY ABERDEEN TEST CENTER
ABERDEEN PROVING GROUND, MD 21005-5059

SEPTEMBER 2009

Prepared for:
U.S. ARMY ENVIRONMENTAL COMMAND
ABERDEEN PROVING GROUND, MD 21010-5401

U.S. ARMY DEVELOPMENTAL TEST COMMAND
ABERDEEN PROVING GROUND, MD 21005-5055

DISPOSITION INSTRUCTIONS

Destroy this document when no longer needed. Do not return to the originator.

The use of trade names in this document does not constitute an official endorsement or approval of the use of such commercial hardware or software. This document may not be cited for purposes of advertisement.

TEDT-AT-SLE

MEMORANDUM FOR RECORD

SUBJECT: Operations Security (OPSEC) Review of Paper/Presentation

1. The attached document entitled "Scoring Record No. 923" dated September 2009 is provided for review for public disclosure in accordance with AR 530-1 as supplemented. The document is proposed for public release via the internet.

2. I, the undersigned, am aware of the intelligence interest in open source publications and in the subject matter of the information I have reviewed for intelligence purposes. I certify that I have sufficient technical expertise in the subject matter of this document and that, to the best of my knowledge, the net benefit of this public release outweighs the potential damage to the essential secrecy of all related ATC, DTC, ATEC, Army or other DOD programs of which I am aware.

J. Stephen McClung Jr.		September 2009
NAME (Printed)	SIGNATURE	DATE

CONCURRENCE:	NAME (Printed)	SIGNATURE	DATE
Program Mgr/Customer (If not ATC owned technology)	Dennis Teefy		30 Oct 09
Directorate Director	Charles Valz		11/2/09
Directorate OPSEC QC and Team Leader	William Burch		23 Sep 2009
ATC OPSEC Officer/ Security Manager	Jenell Bigham		4 Nov 09
Public Affairs Specialist	Juan Melendez		11-5-09
Technical Director, ATC	John R. Wallace		10 Nov 09

(Return to ATC PAO for further processing)

DTC GO/SES	N/A		

Encl
as

REPORT DOCUMENTATION PAGE						*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
September 2009	Final	16 and 17 March 2009

4. TITLE AND SUBTITLE
STANDARDIZED UXO TECHNOLOGY DEMONSTRATION SITE SCORING RECORD NO. 923

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
McClung, Stephen J.

5d. PROJECT NUMBER
8-CO-160-UXO-021

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Army Developmental Test Command
400 Colleran Road
Aberdeen Proving Ground, MD 21005-5059

8. PERFORMING ORGANIZATION REPORT NUMBER
ATC-10071

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Environmental Command
ATTN: IMAE-RTA
314 Longs Corner Road
Aberdeen Proving Ground, MD 21010-5401

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
Same as Item 8

12. DISTRIBUTION/AVAILABILITY STATEMENT
Distribution unlimited.

13. SUPPLEMENTARY NOTES
None

14. ABSTRACT

This scoring record documents the efforts of the Environmental Security Technology Certification Program. to detect and discriminate inert unexploded ordnance (UXO) utilizing the APG Standardized UXO Technology Demonstration Site calibration lanes and open field sites. This Scoring Record was coordinated by J. Stephen McClung and the Standardized UXO Technology Demonstration Site Scoring Committee. Organizations on the committee include the U.S. Army Corps of Engineers, the Environmental Security Technology Certification Program, the Strategic Environmental Research and Development Program, the Institute for Defense Analysis, the U.S. Army Environmental Command, and the U.S. Army Aberdeen Test Center.

15. SUBJECT TERMS
Environmental Security Technology Certification Program, APG, Standardized UXO Technology Demonstration Site, open field, indirect fire, EM61, pushcart.

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
Unclassified	Unclassified	Unclassified	SAR		19b. TELEPHONE NUMBER *(Include area code)*

ACKNOWLEDGMENTS

Authors:

Dennis Teefy
J. Stephen McClung Jr.
William Burch
Homeland Defense and Sustainment Division (HDSD)
U.S. Army Aberdeen Test Center (ATC)
Aberdeen Proving Ground (APG)

Rick Fling
Aberdeen Test Support Services (ATSS)
Sverdrup Technology, Inc.
Aberdeen Proving Ground

Christina McClung
Aberdeen Data Services Team (ADST)
Logistics Engineering and Information Technology Company (Log.Sec)/Tri-S
Aberdeen Proving Ground

TABLE OF CONTENTS

SECTION 4. TECHNICAL PERFORMANCE RESULTS

PAGE

SECTION 5. APPENDIXES

SECTION 1. GENERAL INFORMATION

1.1 BACKGROUND

Technologies under development for the detection and discrimination of military munitions (MM) (i.e. unexploded ordnance {UXO} and discarded military munitions {DMM}) require testing so that performance can be characterized. To that end, Standardized Test Sites have been developed at Aberdeen Proving Ground (APG), Maryland, and U.S. Army Yuma Proving Ground (YPG), Arizona. These test sites provide a diversity of geology, climate, terrain, and weather as well as diversity in munitions and clutter. Testing at these sites is independently administered and analyzed by the government for the purposes of characterizing technologies, tracking performance with system development, comparing performance of different systems, and comparing performance in different environments.

The Standardized UXO Technology Demonstration Site Program is a multiagency program spearheaded by the U.S. Army Environmental Command (USAEC). The U.S. Army Aberdeen Test Center (ATC) and the U.S. Army Corps of Engineers Engineering Research and Development Center (ERDC) provide programmatic support. The program is being funded and supported by the Environmental Security Technology Certification Program (ESTCP), the Strategic Environmental Research and Development Program (SERDP), and the U.S. Army Environmental Quality Technology (EQT) Program.

1.2 SCORING OBJECTIVES

The objective in the Standardized UXO Technology Demonstration Site Program is to evaluate the detection and discrimination capabilities of a given technology under various field and soil conditions. Inert munitions and clutter items are positioned in various orientations and depths in the ground.

The evaluation objectives are as follows:

a. To determine detection and discrimination effectiveness under realistic scenarios with various targets, geology, clutter, density, topography, and vegetation.

b. To determine cost, time, and workforce requirements to operate the technology.

c. To determine the demonstrator's ability to analyze survey data in a timely manner and provide prioritized Target Lists with associated confidence levels.

d. To provide independent site management to enable the collection of high quality, ground-truth (GT), geo-referenced data for post-demonstration analysis.

1.2.1 <u>Scoring Methodology</u>

a. The scoring of the demonstrator's performance is conducted in two stages: response stage and discrimination stage. For both stages, the probability of detection (P_d) and the false alarms are reported as receiver-operating characteristic (ROC) curves. False alarms are divided into those anomalies that correspond to emplaced clutter items, measuring the probability of clutter detection (P_{cd}) or the probability of false positive (P_{fp}). Those that do not correspond to any known item are termed background alarms. The background alarms are addressed as either probability of background alarm (P_{ba}) or background alarm rate (BAR).

b. The response stage scoring evaluates the ability of the system to detect emplaced targets without regard to ability to discriminate munitions from other anomaly sources. For the blind grid response stage, the demonstrator provides a target response from each and every grid square along with a threshold below which target responses are deemed insufficient to warrant further investigation. This list is generated with minimal processing and, since a value is provided for every grid square, includes amplitudes both above and below the system noise level. For the open field, the demonstrator provides a list of all anomalies deemed to exceed a demonstrator selected target detection threshold. An item (either munition or clutter) is counted as detected if a demonstrator indicates an anomaly within a specified distance (Halo Radius (R_{halo})) of a ground truth item.

c. The discrimination stage evaluates the demonstrator's ability to correctly identify munitions as such and to reject clutter. For the blind grid discrimination stage, the demonstrator provides the output of the discrimination stage processing for each grid square. For the open field, the demonstrator provides the output of the discrimination stage processing for anomaly reported in the response stage. The values in these lists are prioritized based on the demonstrator's determination that a location is likely to contain munitions. Thus, higher output values are indicative of higher confidence that a munitions item is present at the specified location. For digital signal processing, priority ranking is based on algorithm output. For other discrimination approaches, priority ranking may be based on rule sets or human judgment. The demonstrator also specifies the threshold in the prioritized ranking that provides optimum performance, (i.e., that is expected to retain all detected munitions and reject the maximum amount of clutter).

d. The demonstrator is also scored on efficiency and rejection ratios, which measure the effectiveness of the discrimination stage processing. The goal of discrimination is to retain the greatest number of munitions detections from the anomaly list, while rejecting the maximum number of anomalies arising from nonmunitions items. Efficiency measures the fraction of detected munitions retained after discrimination, while the rejection ratio measures the fraction of false alarms rejected. Both measures are defined relative to the maximum number of munitions detectable by the sensor and its accompanying clutter detection/false positive rate or BAR.

e. Based on configuration of the GT at the standardized sites and the defined scoring methodology, in some cases, there exists the possibility of having anomalies within overlapping halos and/or multiple anomalies within halos. In these cases, the following scoring logic is implemented:

(1) In situations where multiple anomalies exist within a single R_{halo}, the anomaly with the strongest response or highest ranking will be assigned to that particular GT item. If the responses or rankings are equal, then the anomaly closest to the GT item will be assigned to the GT item. Remaining anomalies are retained and scored until all matching is complete.

(2) Anomalies located within any R_{halo} that do not get associated with a particular GT item are excess alarms and will be disregarded.

f. In some cases, groups of closely spaced munitions have overlapping halos. The following scoring logic is implemented (fig. A-1 through A-9):

(1) Overall site scores (i.e., P_d) will consider only isolated munitions and clutter items.

(2) GT items that have overlapping halos (both munitions and clutter) will form a group and groups may form chains.

(3) Groups will have a complex halos composed of the composite halos of all its GT items.

(4) Groups will have three scoring factors: groups found, groups identified, and group coverage. Scores will be based on 1:1 matches of anomalies and GT.

(a) Groups Found (Found): the number of groups that have one or more GT items matched divided by the total number of groups. Demonstrators will be credited with detecting a group if any item within the group is matched to an anomaly in their lists.

(b) Groups Identified (ID): the number of groups that have two or more GT items matched divided by the total number of groups. Demonstrators will be credited with identifying that a group is present if multiple items within the composite halo are matched to anomalies in their lists.

(c) Group Coverage (Coverage): the number of GT items matched within groups divided by the total number of GT items within groups. This metric measures the demonstrator accuracy in determining the number of anomalies within a group. If five items are present and only two anomalies are matched, the demonstrator will score 0.4. If all five are matched, the demonstrator will score 1.0.

(5) Location error will not be reported for groups.

(6) Demonstrators will not be asked to call out groups in their scoring submissions. If multiple anomalies are indicated in a small area, the demonstrator will report all individual anomalies.

(7) Excess alarms within a halo will be disregarded.

g. All scoring factors are generated utilizing the Standardized UXO Probability and Plot Program, version 4.

1.2.2 Scoring Factors

Factors to be measured and evaluated as part of this demonstration include:

a. Response stage ROC curves:

(1) Probability of detection (P_d^{res}).

(2) Probability of clutter detection (P_{cd}).

(3) Background alarm rate (BAR^{res}) or probability of background alarm (P_{ba}^{res}).

b. Discrimination stage ROC curves:

(1) Probability of detection (P_d^{disc}).

(2) Probability of false positive (P_{fp}).

(3) Background alarm rate (BAR^{disc}) or probability of background alarm (P_{ba}^{disc}).

c. Metrics:

(1) Efficiency (E).

(2) False positive rejection rate (R_{fp}).

(3) Background alarm rejection rate (R_{ba}).

d. Other:

(1) Probability of detection by size, depth, and density.

(2) Classification by type (i.e., 20-, 40-, 105mm, etc.).

(3) Location accuracy for single munitions.

(4) Equipment setup, calibration time, and corresponding worker-hour requirements.

(5) Survey time and corresponding worker-hour requirements.

(6) Reacquisition/resurvey time and worker-hour requirements (if any).

(7) Downtime due to system malfunctions and maintenance requirements.

SECTION 2. DEMONSTRATION

2.1 DEMONSTRATOR INFORMATION

2.1.1 Demonstrator Point of Contact (POC) and Address

POC: Mr. Herb Nelson
703-696-8726

Address: Environmental Security Technology Certification Program
901 North Stuart Street, Suite 303
Arlington, VA 22203

2.1.2 System Description (provided by demonstrator)

The EM61 MK2, the most frequently used geophysical sensor for munitions response projects, is a time-domain electromagnetic induction sensor that samples the decay of the induced field in four time gates after turn off of the primary field. The sensor can be operated in differential mode, with three of the available time gates sampling the decay on the lower receive coil and one sampling an upper, differential coil, or 4 mode with all four gates sampling the lower coil. For this demonstration, the 4 mode will be used.

The sensor is deployed on its standard wheels, resulting in a sensor-to-ground offset of approximately 42 cm. Sensor locations are provided by an attached Global Positioning System (GPS) whose output is logged in the system data logger as the sensor data are being logged. Electromagnetic interference (EMI) data are logged at 10 Hz, which when combined with the nominal 1-m/s survey speed, results in a down-track sample spacing of 10 cm. Survey line spacing will be 50 cm. A series of guidance strings will be placed on the field with a spacing of 1.5 meters. The sensor operator will survey lines to the left, directly over, and to the right of each string. The EM61 MK2 is shown in Figure 1.

Figure 1. NRL, EM61 MK2/pushcart.

Support equipment required: Overnight storage for the sensor that protects it from the elements and access to electrical power for battery charging is required. This and workspace for the data quality control analyst can be located in the building on the site.

Frequency and radio utilization: We are licensed for the following frequencies for GPS corrections (currently using 464.6250 MHz but that can be changed to any of the licensed frequencies):

461.0250 MHz	462.1250 MHz	464.5000 MHz	464.6250 MHz	464.7250 MHz
461.0750 MHz	462.3750 MHz	464.5500 MHz	464.6500 MHz	464.7500 MHz
461.1000 MHz	462.4000 MHz	464.6000 MHz	464.7000 MHz	

2.1.3 Data Processing Description (provided by demonstrator)

Target selection criteria: Targets for this demonstration will be chosen based on the signal expected for each item in the indirect fire area at the depth of interest. NRL have previously determined the minimum sensor response versus depth for each of the three items of interest in this demonstration. An example of such a curve for the 60-mm mortar is shown in Figure 2. The blue line in the figure represents the calculated response of the mortar in its least favorable orientation as a function of depth along with some validating measurements in a test pit.

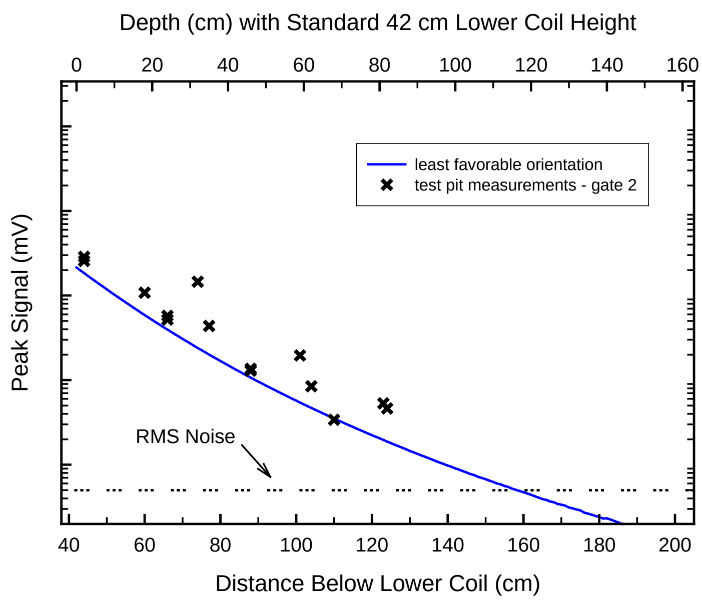


Figure 2. Response of an EM61 MK2 to a 60-mm mortar in its least favorable orientation as a function of the item's depth. The blue curve represents the calculated response, and the symbols are the result of validation measurements in a test pit.

The curves for the three items in the indirect fire area will be confirmed using data from the calibration lanes and the anomaly selection threshold set as the smallest response expected from the three items with a safety margin determined from the calibration lane data.

Parameter estimation:

a. Which characteristics will be extracted from each detected item and input to the discrimination algorithm (e.g., depth, size, polarizability coefficients, fit quality, etc.)? A combination of the polarizability coefficients and dipole fit error.

b. Why have these characteristics been chosen and not others (e.g. empirical evidence of their ability to help discriminate, inclusion in a theoretical tradition, etc.)? The size of the target and its shape can be estimated from the polarizability coefficients and in conjunction with the fit error have previously been determined to help discriminate clutter from UXO.

c. How are these characteristics estimated (e.g., least-mean-squares fit to a dipole model, etc.), to include the equations that are used for parameter estimation? This discrimination approach uses a model-based estimation procedure to determine whether an unknown target is likely to be a UXO item. The EMI signal is a linear function of the flux through the receiving coil. In this model, the flux is assumed to originate from an induced dipole moment at the target location given by the following:

$$\mathbf{m} = \mathrm{UBU}^{\mathrm{T}}\mathbf{H}_0$$

where H_0 is the peak primary field at the target, U is the transformation matrix between the coordinate directions and the principal axes of the target, and B is an empirically determined, effective magnetic polarizability matrix. For an arbitrary compact object, this matrix can be diagonalized about three primary body axes and written as follows:

$$\mathrm{B} = \begin{bmatrix} \beta_{\mathrm{X}} & 0 & 0 \\ 0 & \beta_{\mathrm{Y}} & 0 \\ 0 & 0 & \beta_{\mathrm{Z}} \end{bmatrix}$$

The relative magnitudes of the polarizability coefficients (β's) are determined by the size, shape, and composition of the object and the transmit waveform and time gate or frequency. The transformation matrix contains the angular information about the orientation of these body axes.

The inversion returns the following model parameters: location (X_0, Y_0, Z_0), polarizability coefficients (β_x, β_y, β_z), and orientation angles (*yaw*, *pitch*, *roll*). A nonlinear Marquardt-Levenberg search will be used to determine all nine parameters.

The dipole fit error is defined as $\varepsilon = \sqrt{1 - r^2}$ x 100, where r^2 is the squared correlation coefficient between the best model fit and measured anomaly data.

d. What tunable parameters (if any) are used in the characterization process? (e.g., thresholds on background noise, etc.)? There are no tunable parameters used in the characterization process within UX-Analyze.

Classification:

a. What algorithm is used for discrimination (e.g., multi-layer perception, support vector machine, etc.)? For classification, either the generalized likelihood ratio test (GLRT) or rule-based criterion is used.

b. Why is this algorithm used and not others? The GLRT is the only statistical classifier available in UX-Analyze. It was selected because of its overall performance and ease of use.

c. Which parameters are considered as possible inputs to the algorithm? Inputs to the classification algorithm will consist of combinations of the principal axis polarizabilities. The dipole fit error may also be an input.

d. What are the outputs of the algorithm (probabilities, confidence levels)? The GLRT returns a confidence metric for each anomaly that indicates the likelihood that the subject anomaly is similar to the training items labeled as a target of interest.

e. How is the threshold set to decide where the munitions/non-munitions line lies in the discrimination process? A threshold in parameter space (either in units of GLRT probabilities or attribute space) will be established using the labeled data that separates known TOI from non-TOI.

Training:

a. Which tunable parameters have final values that are optimized over a training set of data and which have values that are set according to geophysical knowledge (i.e., intuition, experience, common sense)? Polarizability coefficients are optimized over a training set of data.

(1) For those tunable parameters with final values set according to geophysical knowledge:

(a) What is the reasoning behind choosing these particular values? N/A.

(b) Why were the final values not optimized over a training set of data? N/A.

(2) For those tunable parameters with final values optimized over the training set data:

(a) What training data are used (e.g., all data, a randomly chosen portion of data, etc.)? All training data on UXO and clutter acquired over the calibration grid and any data taken over the test pit at APG.

(b) What error metric is minimized during training (e.g., mean squared error, etc.)? Standard deviation.

(c) What learning rule is used during training (e.g., gradient descent, etc.)? N/A.

(d) What criterion is used to stop training (e.g., number of iterations exceeds threshold, good generalization over validation set of data, etc.)? Limits of training data.

(e) Are all tunable parameters optimized at once or in sequence (in sequence = parameter 1 is held constant at some common sense values while parameter 2 is optimized, and then parameter 2 is held constant at its optimized value while parameter 1 is optimized)? All at once.

b. What are the final values of all tunable parameters for the characterization process? Best threshold setting.

2.1.4 Data Submission Format

Data were submitted for scoring in accordance with data submission protocols outlined on the USAEC Web site www.uxotestsites.org. These submitted data are not included in this report in order to protect GT information.

2.1.5 Demonstrator Quality Assurance (QA) and Quality Control (QC) (provided by demonstrator)

Overview of Quality Control (QC): Two items need to be checked daily to ensure adequate system performance: geophysical sensor response and reliability of GPS positions. Before beginning survey work each day, the performance of the EM61-MK2 is measured (after a 10- to 15-min warmup) by presenting a standard target to the sensor. The resulting signal is checked against standard values.

The standard procedure is to accept only data with a GPS fix quality of 3 (real-time kinematic (RTK) fixed) or 2 (RTK float) and a precision dilution of precision (PDOP) of 4 or less. Before arriving at the site each day, standard GPS planning software is used to calculate the number of satellites that will be visible to the receivers and the PDOP achievable minute-by-minute throughout the day. This allows short breaks during periods of poor satellite availability to be planned and keeps data that will have to be discarded later from inadvertently being taken. Another important feature of this GPS planning is the ability to take into account areas of restricted sky view (such as along the tree line at one edge of the APG site). Based on past experience, there is usually a brief period each day, from 20 to 30 minutes, when good fixes can be obtained in even the most difficult environments. With planning, the system can be poised by the tree line ready to take data when the appropriate satellite alignment occurs.

Overview of Quality Assurance (QA): At the end of each 1-hour survey session, all survey data are transferred to the field data analyst for preliminary data quality checks. This process involves plotting the actual survey path as logged in the GPS files (color-coded by GPS fix quality) to ensure that GPS data of sufficient quality were obtained during the survey. Following this, the sensor file is examined for completeness and consistency. It is at this stage that any sensor malfunctions, drifts, etc., are flagged and reported to the field crew for correction. The final task for the field analyst is to calculate a position for each sensor reading and apply it to the reading. The mapped data files are then ready for analysis either in the field or at a later time.

2.1.6 Additional Records

The following record(s) by this vendor can be accessed via the Internet as MicroSoft Word documents at www.uxotestsites.org.

2.2 APG SITE INFORMATION

2.2.1 Location

The APG Standardized Test Site is located within a secured range area of the Aberdeen Area. The Aberdeen Area of APG is located approximately 30 miles northeast of Baltimore at the northern end of the Chesapeake Bay. The Standardized Test Site encompasses 17 acres of upland and lowland flats, woods, and wetlands.

2.2.2 Soil Type

According to the soils survey conducted for the entire area of APG in 1998, the test site consists primarily of Elkton Series type soil (ref 2). The Elkton Series consist of very deep, slowly permeable, poorly drained soils. These soils formed in silty aeolin sediments and the underlying loamy alluvial and marine sediments. They are on upland and lowland flats and in depressions of the Mid-Atlantic Coastal Plain. Slopes range from 0 to 2 percent.

ERDC conducted a site-specific analysis in May 2002 (ref 3). The results basically matched the soil survey mentioned above. Seventy percent of the samples taken were classified as silty loam. The majority (77 percent) of the soil samples had a measured water content between 15 and 30 percent with the water content decreasing slightly with depth.

For more details concerning the soil properties at the APG test site, go to www.uxotestsites.org on the Web to view the entire soils description report.

2.2.3 Test Areas

A description of the test site areas at APG is presented in Table 1. A test site layout is shown in Figure 3.

TABLE 1. TEST SITE AREAS

Area	Description
Calibration lanes	Contains 14 standard munitions items buried in six positions, with representation of clutter, at various angles and depths to allow demonstrators to calibrate their equipment.
Blind grid	Contains 400 grid cells in a 0.5-acre site. The center of each grid cell contains either munitions, clutter, or nothing.
Open field	A 10-acre site composed of generally open and flat terrain with minimal clutter and minor navigational obstacles. Vegetation height varies from 15 to 25 cm. This area is subdivided into four subareas (legacy, direct fire, indirect fire, and challenge).
	• *Open field (legacy)* The legacy subarea contains the same wide variety of randomly-placed munitions that were present in the open field prior to the January 2008 general reconfiguration of the site.
	• *Open field (direct fire)* The direct fire subarea contains only three munition types that could be typically found at an impact area of a direct fire weapons range. Munitions and clutter are placed in a pattern typical for these munitions.
	• *Open field (indirect fire)* The indirect fire subarea contains only three munition types that could be typically found at an impact area of an indirect fire weapons range. Munitions and clutter are placed in a pattern typical for these munitions.
	• *Open field (challenge)* The challenge subarea is easily reconfigurable used to meet the specific needs and requirements of the demonstrator or the program sponsor. Any results from this area will not be reported in the standardized scoring record.
Woods	1.34-acre area consisting of cleared woods (tree removal with only stumps remaining), partially cleared woods (including all underbrush and fallen trees), and virgin woods (i.e., woods in natural state with all trees, underbrush, and fallen trees left in place).
Moguls	1.30-acre area consisting of two areas (the rectangular or driving portion of the course and the triangular section with more difficult, nondrivable terrain). A series of craters (as deep as 0.91 m) and mounds (as high as 0.91 m) encompass this section.

2.2.4 STANDARD AND NONSTANDARD INERT MUNITIONS TARGETS

The standard and nonstandard munitions items emplaced in the test areas are presented in Table 2. Standardized targets are members of a set of specific munitions items that have identical properties to all other items in the set (caliber, configuration, size, weight, aspect ratio, material, filler, magnetic remanence, and nomenclature). Nonstandard targets are inert munitions items having properties that differ from those in the set of standardized items.

2.3 ATC SURVEY COMMENTS

Testing of this NRL system was done for the sole purpose of evaluating the change in technical performance when there was a change in survey line spacing. The survey was conducted at a line spacing of 0.5 meters and then a subset of the data gathered was prepared to represent a survey with line spacing of 1.0 meters. Both the complete data set at 0.5-meter line spacing and the data subset at 1.0-meter spacing were submitted for scoring under standard procedures. Both results are shown in this scoring record so they can be directly compared. However, when comparing scoring results from this scoring record with the results from other scoring records only the 0.5-meter line spacing results should be compared with other system results.

TABLE 2. INERT MUNITIONS TARGETS

Item	Munition Type	Calibration Lanes	Blind Grid	Open Field Direct Fire	Open field Indirect Fire	Open Field Legacy	Moguls	Woods
20-mm Projectile M55	S	X				X	X	X
25-mm Projectile M794	S	X	X	X				
37-mm Projectile M47	S	X	X	X				
40-mm Projectile MKII Bodies	S	X				X	X	X
BDU-28 Submunition	S	X				X	X	X
BLU-26 Submunition	S	X				X	X	X
M42 Submunition	S	X				X	X	X
57-mm Projectile APC M86	S	X				X	X	X
60-mm Mortar M49A3	S	X	X		X			
2.75-in. Rocket M230	S	X				X	X	X
81-mm Mortar M374	S	X	X		X	X	X	X
105-mm HEAT Rounds M456	S					X	X	X
105-mm HEAT Round M490	S	X	X	X				
105-mm Projectile M60	S	X	X		X	X	X	X
155-mm Projectile M483A1	S	X				X	X	X
20-mm Projectile M55	NS					X	X	X
20-mm Projectile M97	NS					X	X	X
40-mm Projectile M813	NS					X	X	X
60-mm Mortar (JPG)	NS					X	X	X
60-mm Mortar M49	NS					X	X	X
2.75-in. Rocket M230	NS					X	X	X
2.75-in. Rocket XM229	NS					X	X	X
81-mm Mortar (JPG)	NS					X	X	X
81-mm Mortar M374	NS					X	X	X
105-mm Projectile M60	NS					X	X	X
155-mm Projectile M483A	NS					X	X	X

S = Standard munition.
NS = Nonstandard munition.
JPG = Jefferson Proving Ground.
HEAT = high-explosive antitank.

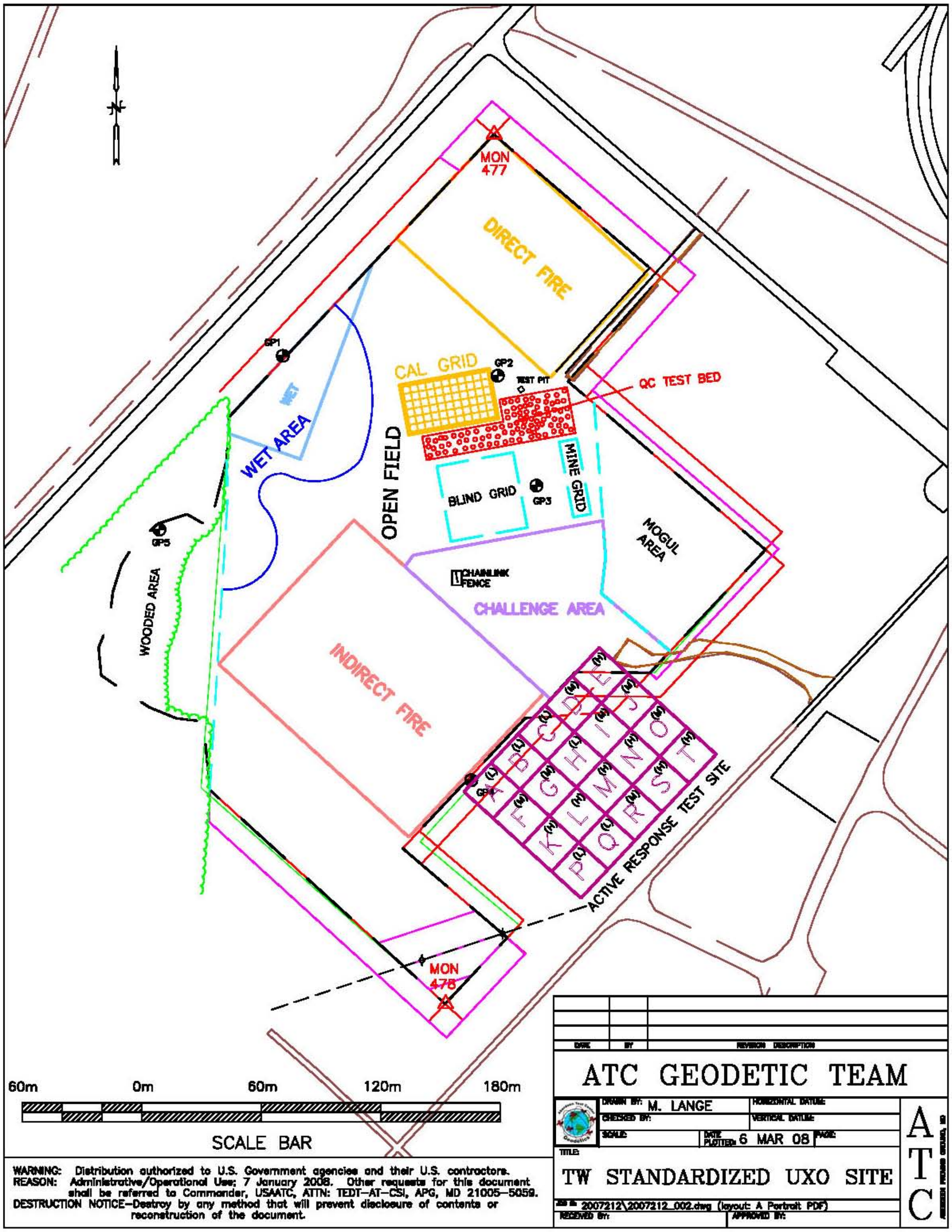


Figure 3. Test site layout.

SECTION 3. FIELD DATA

3.1 DATE OF FIELD ACTIVITIES (16 and 17 March 2009)

3.2 AREAS TESTED/NUMBER OF HOURS

Areas tested and total numbers of hours operated at each site are presented in Table 3.

TABLE 3. AREAS TESTED AND NUMBER OF HOURS

Area	Number of Hours
Calibration lanes	1.66
Blind grid	--
Open field	10.42
Woods	--
Mogul	--
Mine grid	--

Note: Table 3 represents the total time spent in each area.

3.3 TEST CONDITIONS

3.3.1 Weather Conditions

An APG weather station located approximately 1 mile west of the test site was used to record average temperature and precipitation on a half hour basis for each day of operation. The temperatures presented in Table 4 represent the average temperature during field operations from 0700 to 1700 hours, while precipitation data represents a daily total amount of rainfall. Hourly weather logs used to generate this summary are provided in Appendix B.

TABLE 4. TEMPERATURE/PRECIPITATION DATA SUMMARY

Date, 09	Average Temperature, ^{o}F	Total Daily Precipitation, in.
16 Mar	46.2	0.05
17 Mar	47.2	0.04

3.3.2 Field Conditions

NRL surveyed the calibration grid and indirect fire area of the open field. Numerous puddles and wet areas from rain prior to testing were present in the indirect fire area.

3.3.3 Soil Moisture

Three soil probes were placed at various locations within the site to capture soil moisture data: blind grid, calibration, open field, and wooded areas. Measurements were collected in percent moisture and were taken twice daily (morning and afternoon) from five different soil depths (1 to 6 in., 6 to 12 in., 12 to 24 in., 24 to 36 in., and 36 to 48 in.) from each probe. Soil moisture logs are provided in Appendix C.

3.4 FIELD ACTIVITIES

3.4.1 Setup/Mobilization

These activities included initial mobilization and daily equipment preparation and breakdown. A six-person crew took 1 hour and 10 minutes to perform the initial setup and mobilization. There were 25 minutes of daily equipment preparation, and end of the day equipment breakdown lasted 15 minutes.

3.4.2 Calibration

NRL spent a total of 1 hour and 40 minutes in the calibration lanes, of which 70 minutes were spent collecting data. Two other calibration activities occurred during the survey of the indirect fire, totaling 15 minutes.

3.4.3 Downtime Occasions

Occasions of downtime are grouped into five categories: equipment/data checks or equipment maintenance, equipment failure and repair, weather, demonstration site issues, or breaks/lunch. All downtime is included for the purposes of calculating labor requirements (section 5) except for downtime due to demonstration site issues. Demonstration site issues, while noted in the daily log, are considered nonchargeable downtime for the purposes of calculating labor costs and are not discussed. Breaks and lunches are discussed in this section and billed to the total site survey area.

3.4.3.1 Equipment/data checks, maintenance. Equipment data checks and maintenance activities accounted for 30 minutes of site usage time. These activities included changing out batteries and performing routine data checks to ensure the data were being properly recorded/collected. NRL spent no time for breaks and lunches.

3.4.3.2 Equipment failure or repair. Fifteen minutes were needed to resolve equipment failures that occurred while surveying. A poor satellite GPS reading resulted in a 15-minute delay of survey time. It passed, and no further action was required.

3.4.3.3 Weather. No weather delays occurred during the survey.

3.4.4 Data Collection

TABLE 5. TOTAL TIME NRL, SPENT PER AREA

AREA	Time, hr/min
Blind grid	--
Open field	9 hours
Legacy	--
Direct fire	--
Indirect fire	9 hours
Challenge	--
Wooded	--
Mine Grid	--
Moguls	--

Note: Table 5 represents the total time spent in each area collecting data.

3.4.5 Demobilization

The NRL survey crew conducted a demonstration of the calibration grid and indirect fire. Demobilization occurred on 17 March 2009. On that day, it took the crew 1 hour and 5 minutes to break down and pack up their equipment.

3.5 PROCESSING TIME

NRL submitted the raw data from the demonstration activities on the last day of the demonstration, as required. The scoring submittal data were provided 21 May 2009.

3.6 DEMONSTRATOR'S FIELD PERSONNEL

Herb Nelson
Anne Andrews
Jeff Fairbanks
Katherine Kaye
Dan Stinehurst
Glenn Harbaugh

3.7 DEMONSTRATOR'S FIELD SURVEYING METHOD

NRL collected the data in a linear fashion, using a line spacing of 1.5 meters.

3.8 SUMMARY OF DAILY LOGS

Daily logs capture all field activities during this demonstration and are provided in Appendix D. Activities pertinent to this specific demonstration are indicated in highlighted text.

SECTION 4.   TECHNICAL PERFORMANCE RESULTS

4.1   ROC CURVES USING ALL MUNITIONS CATEGORIES

The probability of detection for the response stage (P_d^{res}) and the discrimination stage (P_d^{disc}) versus their respective probability of clutter detection or probability of false positive within each area are shown in Figures 4, 5, 8, and 9. The probabilities plotted against their respective background alarm rate within each area are shown in Figures 6, 7, 10 through 11. Both figures use horizontal lines to illustrate the performance of the demonstrator at two demonstrator-specified points: at the system noise level for the response stage, representing the point below which targets are not considered detectable, and at the demonstrator's recommended threshold level for the discrimination stage, defining the subset of targets the demonstrator would recommend digging based on discrimination.

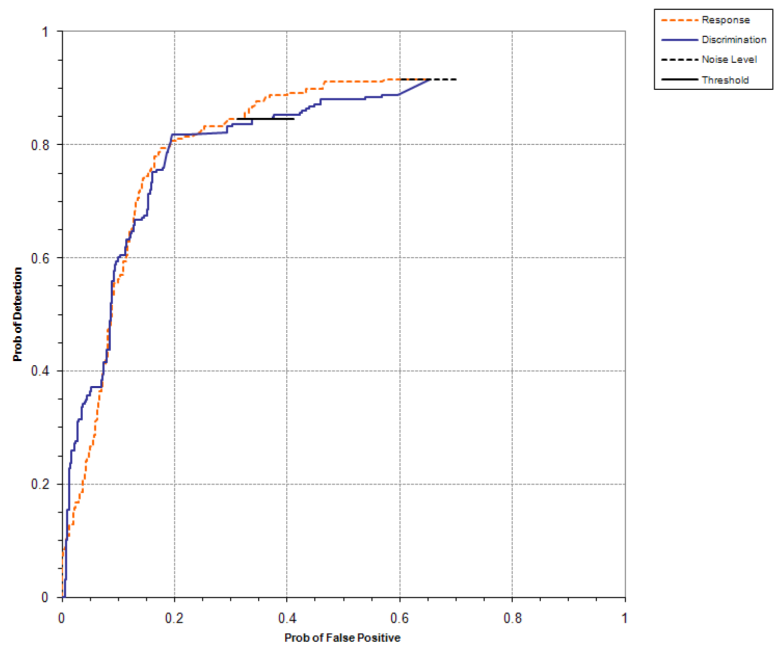


Figure 4.   NRL (0.5-m spacing data set) open field (indirect fire) probability of detection for response and discrimination stages versus their respective probability of false positive.

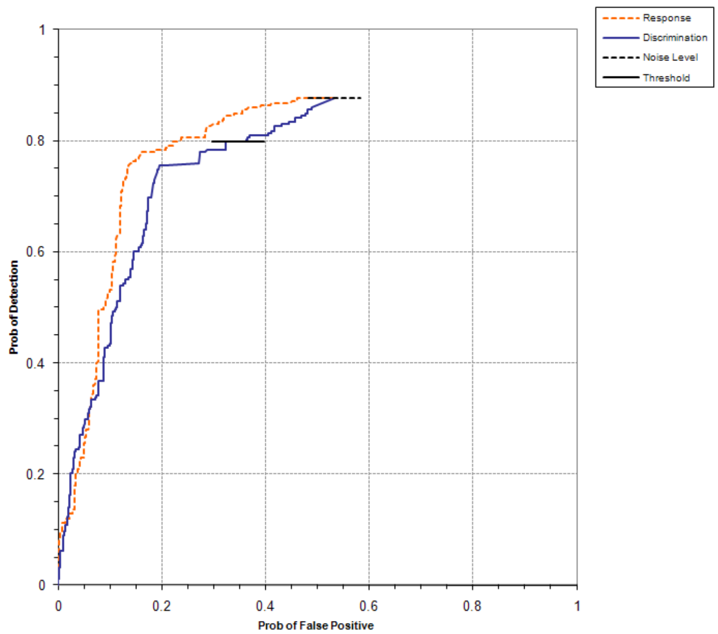


Figure 5.   NRL (1.0-m spacing data set) open field (indirect fire) probability of detection for response and discrimination stages versus their respective probability of false positive.

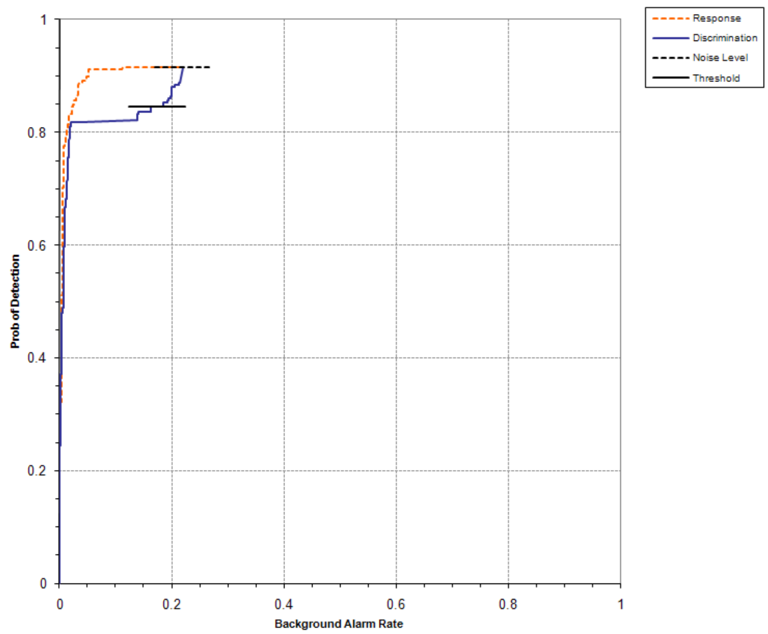


Figure 6.   NRL (0.5-m spacing data set) open field (indirect fire) probability of detection for response and discrimination stages versus their respective background alarm rate.

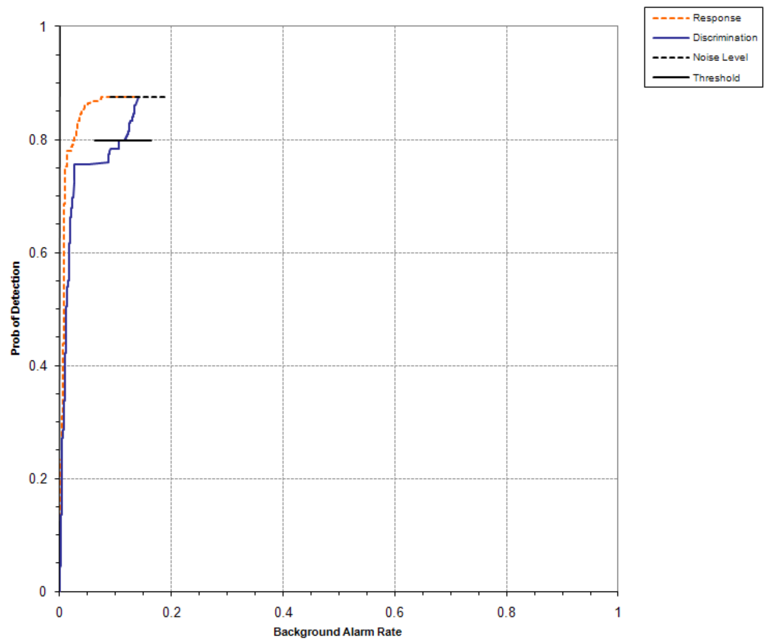


Figure 7. NRL (1.0-m spacing data set) open field (indirect fire) probability of detection for response and discrimination stages versus their respective background alarm rate.

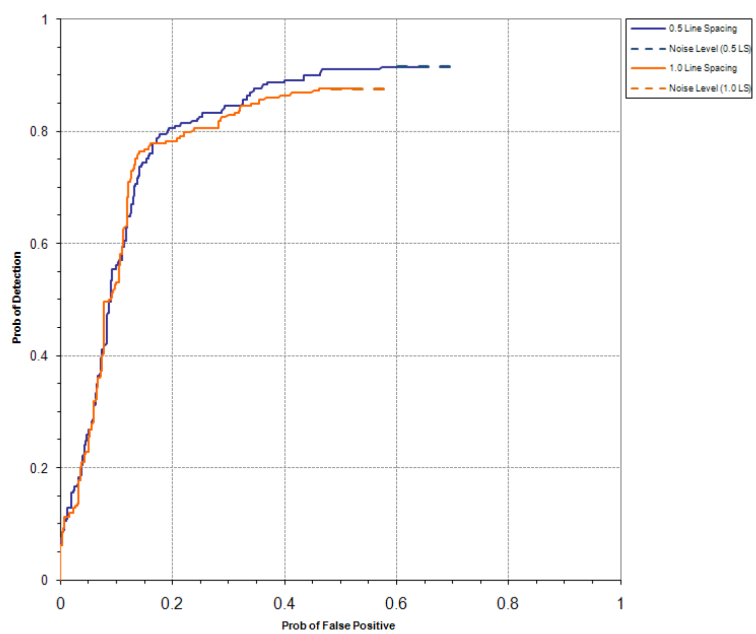


Figure 8. NRL (both data sets) open field (indirect fire) probability of detection for response stage versus their respective probability of false positive.

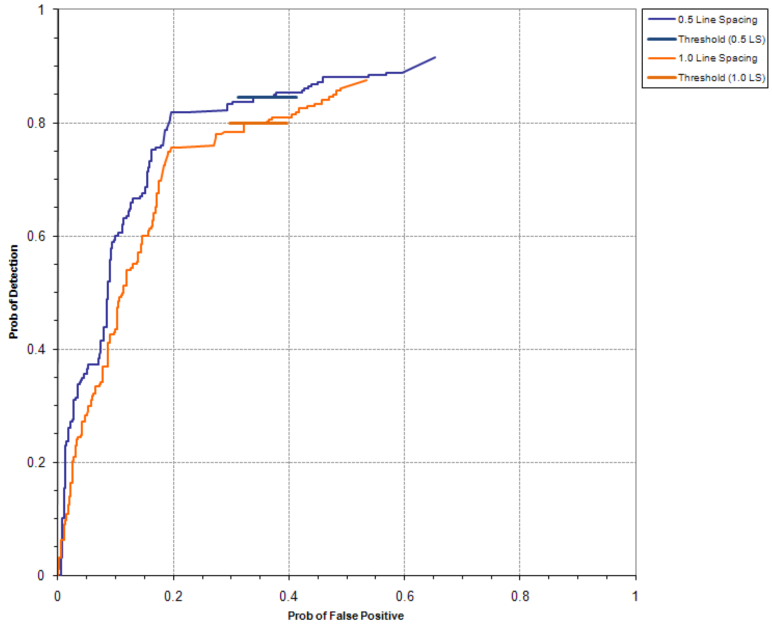


Figure 9. NRL (both data sets) open field (indirect fire) probability of detection for discrimination stage versus their respective probability of false positive.

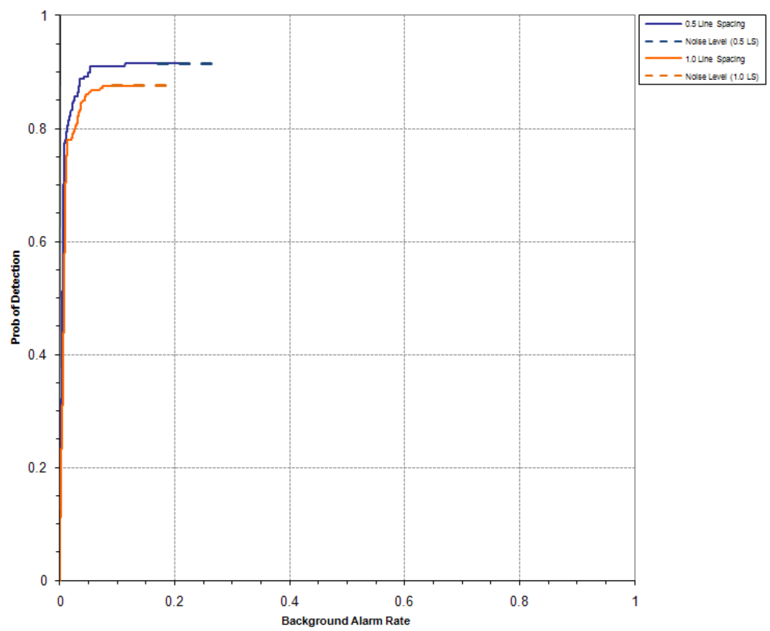


Figure 10. NRL (both data sets) open field (indirect fire) probability of detection for response stage versus their respective background alarm rate.

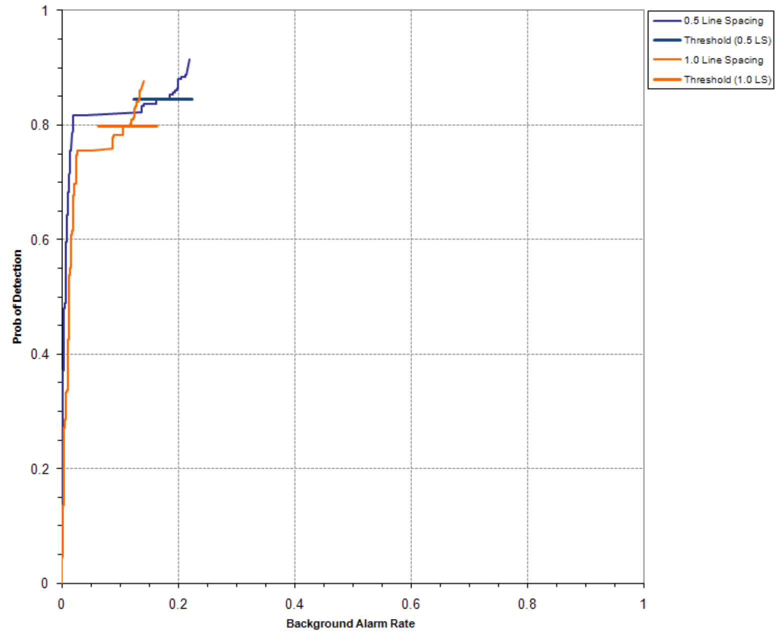


Figure 11. NRL (both data sets) open field (indirect fire) probability of detection for discrimination stage versus their respective background alarm rate.

4.2 PERFORMANCE SUMMARIES

Results for each of the testing areas are presented in Table 6a and 6b. The response stage results are derived from the list of anomalies above the demonstrator-provided noise level. The results for the discrimination stage are derived from the demonstrator's recommended threshold for optimizing munitions related cleanup by minimizing false alarm digs and maximizing munitions recovery. The lower and upper 90-percent confidence limits on P_d, P_{cd}, and P_{fp} were calculated assuming that the number of detections and false positives are binomially distributed random variables.

TABLE 6a. OPEN FIELD INDIRECT FIRE TEST AREA RESULTS (0.5-m spacing data set)

	Response Stage				Discrimination Stage			
[a]Munitions Scores	***P_d^{res}: by type***				***P_d^{disc}: by type***			
	All Types	**105-mm**	**81-mm**	**60-mm**	**All Types**	**105-mm**	**81-mm**	**60-mm**
	0.94 0.92 0.89	0.96 0.93 0.88	0.94 0.90 0.85	0.95 0.91 0.86	0.87 0.85 0.81	0.83 0.77 0.70	0.91 0.87 0.81	0.94 0.90 0.84
By Density								
High	0.92	0.96	0.91	0.88	0.84	0.73	0.91	0.88
Medium	0.87	0.87	0.86	0.90	0.81	0.73	0.79	0.90
Low	0.95	0.97	0.94	0.94	0.89	0.84	0.91	0.91
By Depth[b]								
0 to 4D	0.97	0.98	0.97	0.95	0.89	0.81	0.91	0.95
4D to 8D	0.86	0.90	0.87	0.71	0.79	0.74	0.85	0.71
8D to 12D	0.80	0.50	0.78	0.92	0.76	0.50	0.78	0.83
Clutter Scores	***P_{cd}***				***P_{fp}***			
By Mass								
By Depth[b]	**All Mass**	**0 to 0.25 kg**	**>0.25 to 1 kg**	**>1 to 8 kg**	**All Mass**	**0 to 0.25 kg**	**>0.25 to 1 kg**	**>1 to 8 kg**
All Depth	0.68 0.65 0.62	0.52	0.78	0.87	0.40 0.36 0.33	0.33	0.39	0.44
0 to 0.15 m	0.64	0.52	0.80	0.88	0.37	0.34	0.42	0.42
0.15 to 0.3 m	0.67	0.56	0.63	0.82	0.29	0.25	0.19	0.41
0.3 to 0.6 m	0.83	0.00	0.83	1.00	0.42	0.00	0.33	0.60
Background Alarm Rates								
	BAR^{res}: 0.22				**BAR^{disc}:** 0.17			
Groups								
Found	0.95				0.80			
Identified	0.00				0.00			
Coverage	0.47				0.40			

[a]The two numbers to the right of the all types munitions result are an upper and lower 90-percent confidence interval for an assumed binomial distribution.
[b]All depths are measured to the center of the object.

TABLE 6b. OPEN FIELD INDIRECT FIRE TEST AREA RESULTS (1.0-m spacing data set)

	Response Stage				Discrimination Stage			
[a]Munitions Scores	***P_d^{res}: by type***				***P_d^{disc}: by type***			
	All Types	**105-mm**	**81-mm**	**60-mm**	**All Types**	**105-mm**	**81-mm**	**60-mm**
	0.90 0.88 0.85	0.93 0.89 0.83	0.90 0.86 0.79	0.93 0.89 0.83	0.83 0.80 0.76	0.82 0.76 0.69	0.84 0.78 0.71	0.90 0.85 0.79
By Density								
High	0.82	0.85	0.82	0.80	0.70	0.58	0.77	0.76
Medium	0.85	0.83	0.86	0.86	0.78	0.80	0.75	0.79
Low	0.94	0.97	0.88	0.97	0.89	0.87	0.82	0.97
By Depth[b]								
0 to 4D	0.95	0.96	0.94	0.95	0.87	0.83	0.86	0.92
4D to 8D	0.77	0.81	0.80	0.64	0.69	0.68	0.72	0.64
8D to 12D	0.76	0.50	0.78	0.83	0.72	0.50	0.78	0.75
Clutter Scores	***P_{cd}***				***P_{fp}***			
By Mass								
By Depth[b]	**All Mass**	**0 to 0.25 kg**	**>0.25 to 1 kg**	**>1 to 8 kg**	**All Mass**	**0 to 0.25 kg**	**>0.25 to 1 kg**	**>1 to 8 kg**
All Depth	0.57 0.53 0.50	0.39	0.66	0.78	0.38 0.35 0.32	0.29	0.42	0.37
0 to 0.15 m	0.53	0.39	0.70	0.83	0.36	0.30	0.45	0.38
0.15 to 0.3 m	0.57	0.50	0.50	0.71	0.27	0.25	0.25	0.29
0.3 to 0.6 m	0.42	0.00	0.17	0.80	0.33	0.00	0.17	0.60
Background Alarm Rates								
	BAR^{res}: 0.14				**BAR^{disc}:** 0.11			
Groups								
Found	0.95				0.78			
Identified	0.00				0.00			
Coverage	0.47				0.38			

[a]The two numbers to the right of the all types munitions result are an upper and lower 90-percent confidence interval for an assumed binomial distribution.
[b]All depths are measured to the center of the object.

4.3 EFFICIENCY, REJECTION RATES, AND TYPE CLASSIFICATION

Efficiency and rejection rates are calculated to quantify the discrimination ability at specific points of interest on the ROC curve: (1) at the point where no decrease in P_d is suffered (i.e., the efficiency is by definition equal to one) and (2) at the operator selected threshold. These values are presented in Tables 7a and 7b.

TABLE 7a. OPEN FIELD (INDIRECT) EFFICIENCY AND REJECTION RATES (0.5-m spacing data set)

	Efficiency (E)	False Positive Rejection Rate	Background Alarm Rejection Rate
At operating point	0.92	0.44	0.21
With no loss of P_d	1.00	0.00	0.00

TABLE 7b. OPEN FIELD (INDIRECT) EFFICIENCY AND REJECTION RATES (1.0-m spacing data set)

	Efficiency (E)	False Positive Rejection Rate	Background Alarm Rejection Rate
At operating point	0.91	0.35	0.21
With no loss of P_d	1.00	0.00	0.00

At the demonstrator's recommended setting, the munitions items that were detected and correctly discriminated were further scored on whether their correct type could be identified (tables 8a and 8b). Correct type examples include 20-mm projectile, 105-mm HEAT projectile, and 2.75-inch Rocket. A list of the standard type declaration required for each munitions item was provided to demonstrators prior to testing. The standard types for the three example items are 20-mmP, 105H, and 2.75-inch.

TABLE 8a. OPEN FIELD INDIRECT FIRE CORRECT TYPE CLASSIFICATION OF TARGETS CORRECTLY DISCRIMINATED AS MUNITIONS (0.5-m spacing data set)

Size	Percentage Correct
60 mm	59%
81 mm	47%
105 mm	70%
Overall	59%

TABLE 8b. OPEN FIELD INDIRECT FIRE CORRECT TYPE CLASSIFICATION OF TARGETS CORRECTLY DISCRIMINATED AS MUNITIONS (1.0-m spacing data set)

Size	Percentage Correct
60 mm	40%
81 mm	45%
105 mm	71%
Overall	52%

4.4 LOCATION ACCURACY

The mean location error and standard deviations are presented in Tables 9a and 9b. These calculations are based on average missed distance for munitions correctly identified during the response stage. Depths are measured from the center of the munitions to the surface. For the blind grid, only depth errors are calculated because (X, Y) positions are known to be the centers of the grid square.

TABLE 9a. OPEN FIELD INDIRECT FIRE MEAN LOCATION ERROR AND STANDARD DEVIATION (0.5-m spacing data set)

	Mean	Standard Deviation
Northing	0.02	0.12
Easting	0.00	0.11
Depth	0.03	0.15

TABLE 9b. OPEN FIELD INDIRECT FIRE MEAN LOCATION ERROR AND STANDARD DEVIATION (1.0-m spacing data set)

	Mean	Standard Deviation
Northing	0.01	0.13
Easting	0.02	0.13
Depth	0.03	0.18

4.5 STATISTICAL COMPARISONS

Statistical Chi-square significance tests were used to compare results between the 0.5-meter line spacing and the 1.0-meter line spacing scenarios. The intent of the comparison is to determine if the feature introduced in each scenario has a degrading effect on the performance of the sensor system. However, any modifications in the UXO sensor system during the test, like changes in the processing or changes in the selection of the operating threshold, will also contribute to performance differences.

The Chi-square test for comparison between ratios was used at a significance level of 0.05 meters to compare 0.5-meter line spacing to 1.0-meter line spacing with regard to P_d^{res}, P_d^{disc}, P_{fp}^{res} and P_{fp}^{disc}, Efficiency and Rejection Rate. These results are presented in Table 10. A detailed explanation and example of the Chi-square application is provided in Appendix A.

TABLE 10. CHI-SQUARE RESULTS - 0.5-METER LINE SPACING VERSUS 1.0-METER LINE SPACING

Metric	60-mm	81-mm	105-mm	Overall
P_d^{res}	Not significant	Not significant	Not significant	Not significant
P_d^{disc}	Not significant	Not significant	Not significant	Not significant
P_{fp}^{res}	-	-	-	Significant
P_{fp}^{disc}	-	-	-	Not significant
Efficiency	-	-	-	Not significant
Rejection rate	-	-	-	Significant

SECTION 5. APPENDIXES

APPENDIX A. TERMS AND DEFINITIONS

GENERAL DEFINITIONS

Anomaly: Location of a system response deemed to warrant further investigation by the demonstrator for consideration as an emplaced munitions item.

Detection: An anomaly location that is within R_{halo} of an emplaced munitions item.

Military Munitions (MM): Specific categories of MM that may pose unique explosive safety risks, including UXO as defined in 10 USC 101(e)(5), DMM as defined in 10 USC 2710(e)(2) and/or munitions constituents (e.g. TNT, RDX) as defined in 10 USC 2710(e)(3) that are present in high enough concentrations to pose an explosive hazard.

Emplaced Munitions: A munitions item buried by the government at a specified location in the test site.

Emplaced Clutter: A clutter item (i.e., nonmunitions item) buried by the government at a specified location in the test site.

R_{halo}: A predetermined radius about an emplaced item (clutter or munitions) within which an anomaly identified by the demonstrator as being of interest is considered to be a detection of that item. For the purpose of this program, a circular halo 0.5 meters in radius is placed around the center of the object for all clutter and munitions items.

Small Munitions: Caliber of munitions less than or equal to 40 mm (includes 20-mm projectile, 25-mm projectile, 37-mm projectile, 40-mm projectile, submunitions BLU-26, BLU-63, and M42).

Medium Munitions: Caliber of munitions greater than 40 mm and less than or equal to 81 mm (includes 57-mm projectile, 60-mm mortar, 2.75-inch rocket, and 81-mm mortar).

Large Munitions: Caliber of munitions greater than 81 mm (includes 105-mm HEAT, 105-mm projectile, and 155-mm projectile).

Group: Two or more adjacent GT items with overlapping halos.

GT: Ground truth

Response Stage Noise Level: The level that represents the signal level below which anomalies are not considered detectable. Demonstrators are required to provide the recommended noise level for the blind grid test area.

Discrimination Stage Threshold: The demonstrator-selected threshold level that is expected to provide optimum performance of the system by retaining all detectable munitions and rejecting the maximum amount of clutter. This level defines the subset of anomalies the demonstrator would recommend digging based on discrimination.

Binomially Distributed Random Variable: A random variable of the type which has only two possible outcomes, say success and failure, is repeated for n independent trials with the probability p of success and the probability $1-p$ of failure being the same for each trial. The number of successes x observed in the n trials is an estimate of p and is considered to be a binomially distributed random variable.

RESPONSE AND DISCRIMINATION STAGE DATA

The scoring of the demonstrator's performance is conducted in two stages: response stage and discrimination stage. For both stages, the probability of detection (P_d) and the false alarms are reported as receiver-operating characteristic (ROC) curves. False alarms are divided into those anomalies that correspond to emplaced clutter items, measuring the probability of clutter detection (P_{cd}) or probability of false positive (P_{fp}). Those that do not correspond to any known item are termed background alarms.

The response stage is a measure of whether the sensor can detect an object of interest. For a channel instrument, this value should be closely related to the amplitude of the signal. The demonstrator must report the response level (threshold) below which target responses are deemed insufficient to warrant further investigation. At this stage, minimal processing may be done. This includes filtering long- and short-scale variations, bias removal, and scaling. This processing should be detailed in the data submission.

For a multichannel instrument, the demonstrator must construct a quantity analogous to amplitude. The demonstrator should consider what combination of channels provides the best test for detecting any object that the sensor can detect. The average amplitude across a set of channels is an example of an acceptable response stage quantity. Other methods may be more appropriate for a given sensor. Again, minimal processing can be done, and the demonstrator should explain how this quantity was constructed in their data submission.

The discrimination stage evaluates the demonstrator's ability to correctly identify munitions as such, and to reject clutter. For the same locations as in the response stage anomaly list, the discrimination stage list contains the output of the algorithms applied in the discrimination-stage processing. This list is prioritized based on the demonstrator's determination that an anomaly location is likely to contain munitions. Thus, higher output values are indicative of higher confidence that a munitions item is present at the specified location. For electronic signal processing, priority ranking is based on algorithm output. For other systems, priority ranking is based on human judgment. The demonstrator also selects the threshold that the demonstrator believes will provide optimum system performance, (i.e., that retains all the detected munitions and rejects the maximum amount of clutter).

Note: The two lists provided by the demonstrator contain identical numbers of potential target locations. They differ only in the priority ranking of the declarations.

GROUP SCORING FACTORS

Based on configuration of the GT at the standardized sites and the defined scoring methodology, there exists munitions groups defined as having overlapping halos. In these cases, the following scoring logic is implemented (fig. A-1 through A-9):

a. Overall site scores (i.e., P_d) will consider only isolated munitions and clutter items.

b. GT items that have overlapping halos (both munitions and clutter) will form a group and groups may form chains.

c. Groups will have a complex halos composed of all the composite halos of all its GT items.

d. Groups will have three scoring factors: groups found groups identified and group coverage. Scores will be based on 1:1 matches of anomalies and GT.

(1) Groups Found (Found): the number of groups that have one or more GT items matched divided by the total number of groups. Demonstrators will be credited with detecting a group if any item within the group is matched to an anomaly in their list.

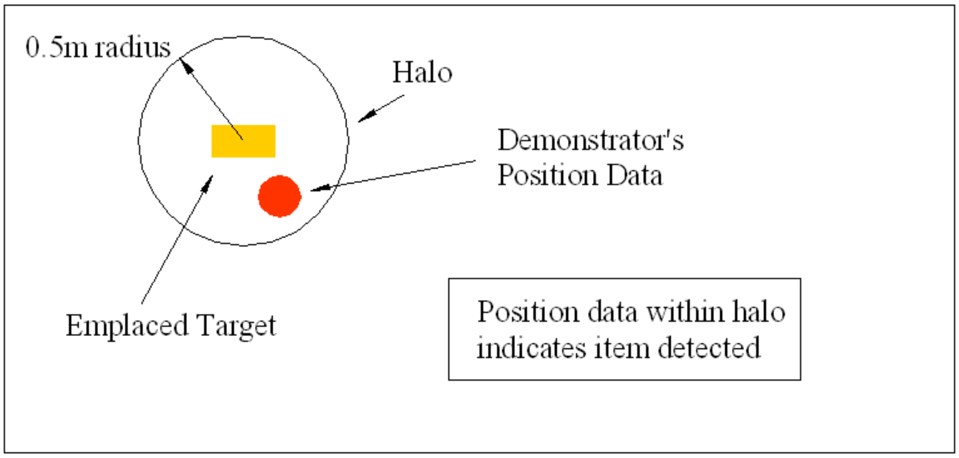
(2) Groups Identified (ID): the number of groups that have two or more GT items matched divided by the total number of groups. Demonstrators will be credited with identifying that a group is present if multiple items within the composite halo are matched to anomalies in their list.

(3) Group Coverage (Coverage): the number of GT items matched within groups divided by the total number of GT items within groups. This metric measures the demonstrator accuracy in determining the number of anomalies within a group. If five items are present and only two anomalies are matched, the demonstrator will score 0.4. If all five are matched the demonstrator will score 1.0.

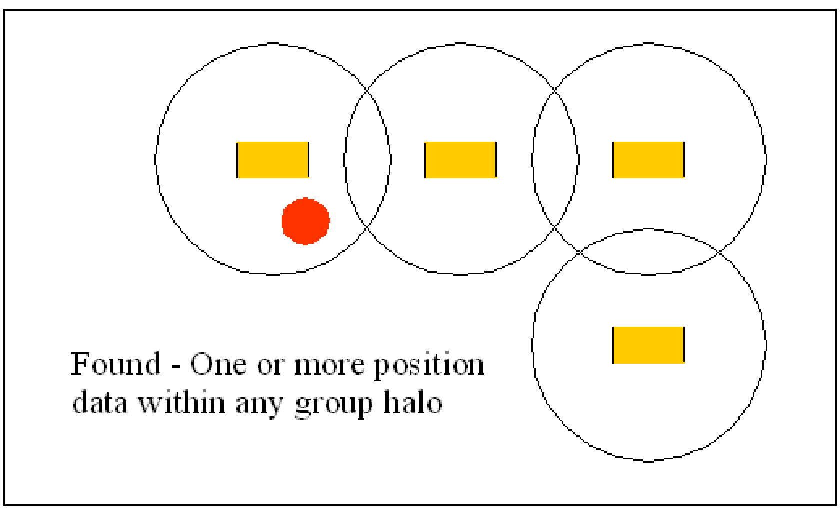
e. Location error will not be reported for groups.

f. Demonstrators will not be asked to call out groups in their scoring submissions. If multiple anomalies are indicated in a small area, the demonstrator will report all individual anomalies.

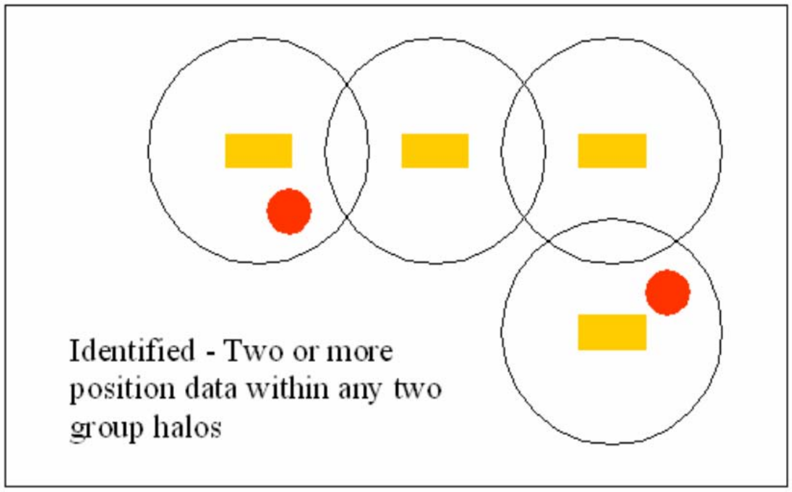
g. Excess alarms within a halo will be disregarded.



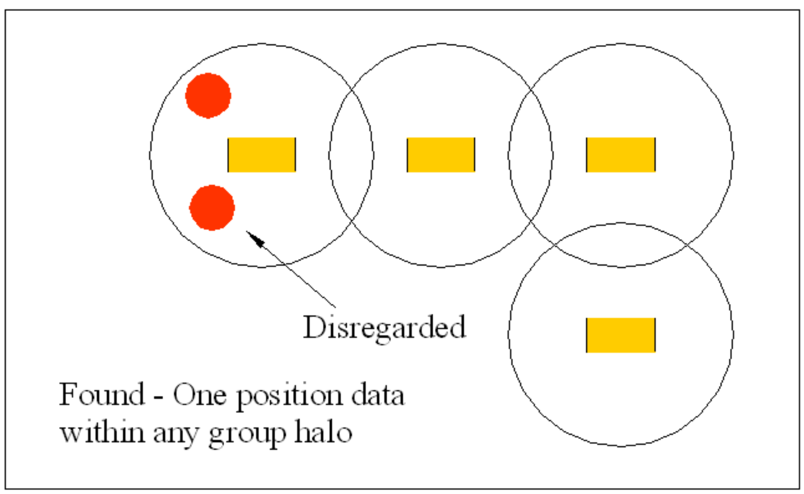
A-1. Example of detected item.



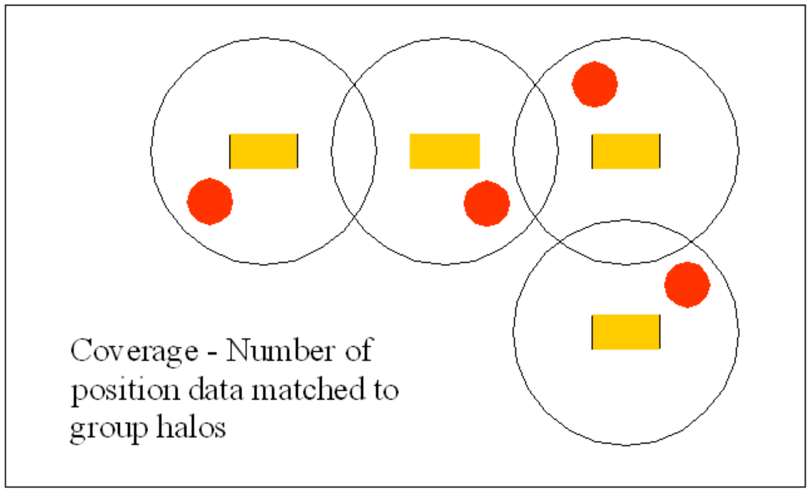
A-2. Example of group found (found).



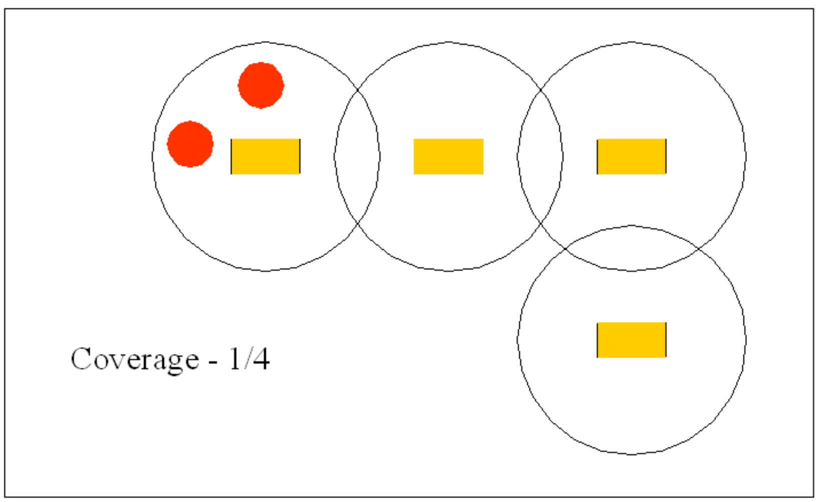
A-3. Example of group identified (ID).



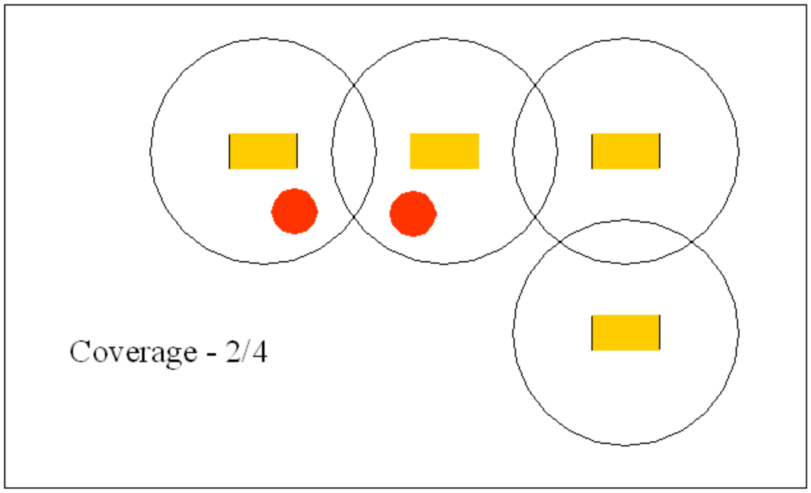
A-4. Example of excess alarms disregarded.



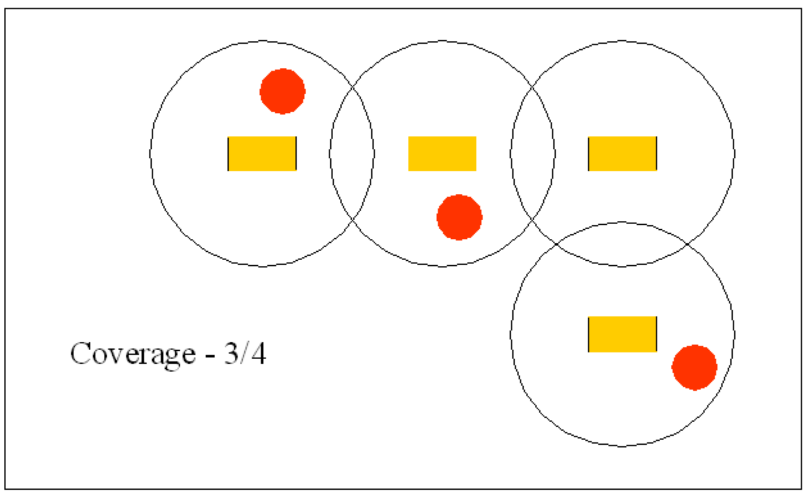
A-5. Example of a group.



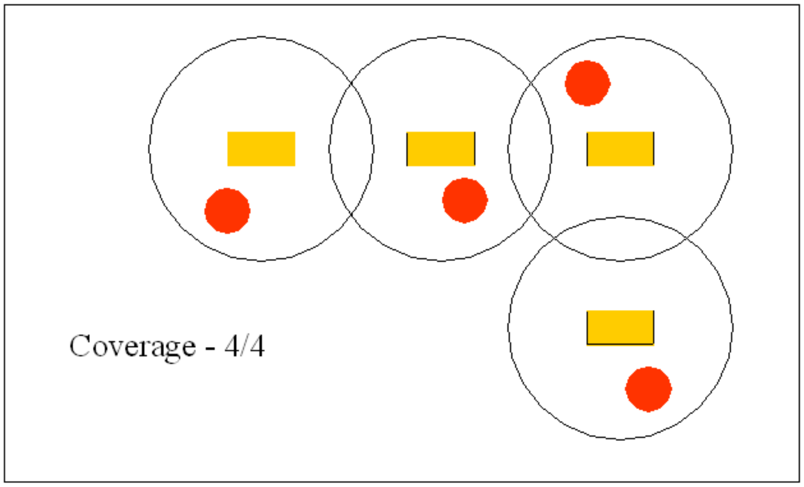
A-6. Example of group (1/4 = 0.25).



A-7. Example of group (2/4 = 0.5).



A-8. Example of group (3/4 = 0.75).



A-9. Example of group (4/4 = 1.0).

RESPONSE STAGE DEFINITIONS

Response Stage Probability of Detection (P_d^{res}): P_d^{res} = (No. of response-stage detections)/(No. of emplaced munitions in the test site).

Response Stage Clutter Detection (cd^{res}): An anomaly location that is within R_{halo} of an emplaced clutter item.

Response Stage Probability of Clutter Detection (P_{cd}^{res}): P_{cd}^{res} = (No. of response-stage clutter detections)/(No. of emplaced clutter items).

Response Stage Background Alarm (ba^{res}): An anomaly in a blind grid cell that contains neither emplaced munitions nor an emplaced clutter item. An anomaly location in the open field or scenarios that is outside R_{halo} of any emplaced munitions or emplaced clutter item.

Response Stage Probability of Background Alarm (P_{ba}^{res}): Blind grid only: P_{ba}^{res} = (No. of response-stage background alarms)/(No. of empty grid locations).

Response Stage Background Alarm Rate (BAR^{res}): Open field any challenge area (including the direct and indirect firing sub areas) only: BAR^{res} = (No. of response-stage background alarms)/(arbitrary constant).

Note that the quantities P_d^{res}, P_{cd}^{res}, P_{ba}^{res}, and BAR^{res} are functions of t^{res}, the threshold applied to the response-stage signal strength. These quantities can therefore be written as $P_d^{res}(t^{res})$, $P_{cd}^{res}(t^{res})$, $P_{ba}^{res}(t^{res})$, and $BAR^{res}(t^{res})$.

DISCRIMINATION STAGE DEFINITIONS

Discrimination: The application of a signal processing algorithm or human judgment to sensor data to discriminate munitions from clutter. Discrimination should identify anomalies that the demonstrator has high confidence correspond to munitions, as well as those that the demonstrator has high confidence correspond to nonmunitions or background returns. The former should be ranked with highest priority and the latter with lowest.

Discrimination Stage Probability of Detection (P_d^{disc}): P_d^{disc} = (No. of discrimination-stage detections)/(No. of emplaced munitions in the test site).

Discrimination Stage False Positive (fp^{disc}): An anomaly location that is within R_{halo} of an emplaced clutter item.

Discrimination Stage Probability of False Positive (P_{fp}^{disc}): P_{fp}^{disc} = (No. of discrimination stage false positives)/(No. of emplaced clutter items).

Discrimination Stage Background Alarm (ba^{disc}): An anomaly in a blind grid cell that contains neither emplaced munitions nor an emplaced clutter item. An anomaly location in the open field or scenarios that is outside R_{halo} of any emplaced munitions or emplaced clutter item.

Discrimination Stage Probability of Background Alarm (P_{ba}^{disc}): P_{ba}^{disc} = (No. of discrimination-stage background alarms)/(No. of empty grid locations).

Discrimination Stage Background Alarm Rate (BAR^{disc}): BAR^{disc} = (No. of discrimination-stage background alarms)/(arbitrary constant).

Note that the quantities P_d^{disc}, P_{fp}^{disc}, P_{ba}^{disc}, and BAR^{disc} are functions of t^{disc}, the threshold applied to the discrimination-stage signal strength. These quantities can therefore be written as $P_d^{disc}(t^{disc})$, $P_{fp}^{disc}(t^{disc})$, $P_{ba}^{disc}(t^{disc})$, and $BAR^{disc}(t^{disc})$.

RECEIVER-OPERATING CHARACERISTIC (ROC) CURVES

ROC curves at both the response and discrimination stages can be constructed based on the above definitions. The ROC curves plot the relationship between P_d versus P_{cd} or P_{fp} and P_d versus BAR or P_{ba} as the threshold applied to the signal strength is varied from its minimum (t_{min}) to its maximum (t_{max}) value.[1] P_d versus P_{fp} and P_d versus BAR being combined into ROC curves is shown in Figure A-10. Note that the “res” and “disc” superscripts have been suppressed from all the variables for clarity.

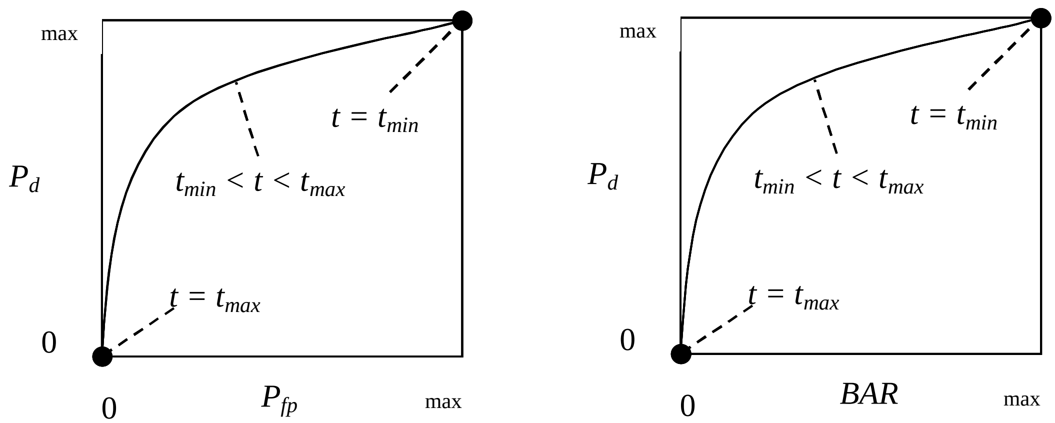


Figure A-10. ROC curves for open field testing. Each curve applies to both the response and discrimination stages.

METRICS TO CHARACTERIZE THE DISCRIMINATION STAGE

The demonstrator is also scored on efficiency and rejection ratio, which measure the effectiveness of the discrimination stage processing. The goal of discrimination is to retain the greatest number of munitions detections from the anomaly list while rejecting the maximum number of anomalies arising from nonmunitions items. The efficiency measures the fraction of detected munitions retained by the discrimination, while the rejection ratio measures the fraction of false alarms rejected. Both measures are defined relative to the entire response list, i.e., the maximum munitions detectable by the sensor and its accompanying clutter detection rate/false positive rate or background alarm rate.

---

[1]Strictly speaking, ROC curves plot the P_d versus P_{ba} over a predetermined and fixed number of detection opportunities (some of the opportunities are located over munitions and others are located over clutter or blank spots). In an open field scenario, each system suppresses its signal strength reports until some bare-minimum signal response is received by the system. Consequently, the open field ROC curves do not have information from low signal-output locations, and, furthermore, different contractors report their signals over a different set of locations on the ground. These ROC curves are thus not true to the strict definition of ROC curves as defined in textbooks on detection theory. Note, however, that the ROC curves obtained in the blind grid test sites are true ROC curves.

Efficiency (E): $E = P_d^{disc}(t^{disc})/P_d^{res}(t_{min}^{res})$: Measures (at a threshold of interest) the degree to which the maximum theoretical detection performance of the sensor system (as determined by the response stage tmin) is preserved after application of discrimination techniques. Efficiency is a number between 0 and 1. An efficiency of 1 implies that all of the munitions initially detected in the response stage were retained at the specified threshold in the discrimination stage, t^{disc}.

False Positive Rejection Rate (R_{fp}): $R_{fp} = 1 - [P_{fp}^{disc}(t^{disc})/P_{cd}^{res}(t_{min}^{res})]$: Measures (at a threshold of interest) the degree to which the sensor system's false positive performance is improved over the maximum false positive performance (as determined by the response stage tmin). The rejection rate is a number between 0 and 1. A rejection rate of 1 implies that all emplaced clutter initially detected in the response stage were correctly rejected at the specified threshold in the discrimination stage.

Background Alarm Rejection Rate (R_{ba}):

Blind grid: $R_{ba} = 1 - [P_{ba}^{disc}(t^{disc})/P_{ba}^{res}(t_{min}^{res})]$.
Open field: $R_{ba} = 1 - [BAR^{disc}(t^{disc})/BAR^{res}(t_{min}^{res})])$.

Measures the degree to which the discrimination stage correctly rejects background alarms initially detected in the response stage. The rejection rate is a number between 0 and 1. A rejection rate of 1 implies that all background alarms initially detected in the response stage were rejected at the specified threshold in the discrimination stage.

CHI-SQUARE COMPARISON

The Chi-square test for differences in probabilities (or 2 by 2 contingency table) is used to analyze two samples drawn from two different populations to see if both populations have the same or different proportions of elements in a certain category. More specifically, two random samples are drawn, one from each population, to test the null hypothesis that the probability of event A (some specified event) is the same for both populations (ref 3).

The test statistic of the 2 by 2 contingency table is the Chi-square distribution with one degree of freedom. When an association between a more challenging terrain feature and relatively degraded performance is sought, a one-sided test is performed. A two-sided 2 by 2 contingency table is used in the Standardized UXO Technology Demonstration Site Program to compare performance between any two areas or subareas when the direction of degradation cannot be predetermined.

For a one-sided test, a significance level of 0.05 is used to set the critical decision limit. It is a critical decision limit because if the test statistic calculated from the data exceeds this value, then the lower proportion tested will be considered significantly less than the greater one (degraded). If the test statistic calculated from the data is less than this value, then no degradation can be said to exist because of the terrain feature introduced.

For a two-sided test, a significance level of 0.10 is used to allow .05 on either side of the decision. It is a critical decision limit because if the test statistic calculated from the data exceeds this value, then the two proportions tested will be considered significantly different. If the test statistic calculated from the data is less than this value, then the two proportions tested will be considered not significantly different.

An exception must be applied when either a 0 or 100 percent success rate occurs in the sample data. The Chi-square test cannot be used in these instances. Instead, Fischer's test is used, and the critical decision limit for one-sided tests is the chosen significance level, which in this case is 0.05. With Fischer's test, if the test statistic is less than the critical value, then the proportions are considered to be significantly different.

An example follows that illustrates Standardized UXO Technology Demonstration Site blind grid results compared to those from the open field legacy. It should be noted that a significant result does not prove a cause-and-effect relationship exists between the two populations of interest; however, it does serve as a tool to indicate that one data set has experienced a degradation or change in system performance at a large enough level than can be accounted for merely by chance or random variation. Note also that a result that is not significant indicates that there is not enough evidence to declare that anything more than chance or random variation within the same population is at work between the two data sets being compared.

Demonstrator X achieves the following overall results after surveying the blind grid and open field (legacy) using the same system (results indicate the number of munitions detected divided by the number of munitions emplaced):

	Blind grid	Open field
P_d^{res}	100/100 = 1.0	8/10 = .80

P_d^{res}: BLIND GRID versus OPEN FIELD (legacy). Using the example data above to compare probabilities of detection in the response stage, all 100 munitions out of 100 emplaced munitions items were detected in the blind grid while 8 munitions out of 10 emplaced were detected in the open field. Fischer's test must be used since a 100 percent success rate occurs in the data. Fischer's test uses the four input values to calculate a test statistic of 0.0075 that is compared against the critical value of 0.05. Since the test statistic is less than the critical value, the smaller response stage detection rate (0.80) is considered to be significantly less at the 0.05 level of significance. While a significant result does not prove a cause-and-effect relationship exists between the change in survey area and degradation in performance, it does indicate that the detection ability of demonstrator X's system seems to have been degraded in the open field relative to results from the blind grid using the same system. This is an example of a one-sided Chi-squared test.

APPENDIX B. DAILY WEATHER LOGS

Date, 09	Time, EST	Avg. Temp, °F	Total Precip., in.
16 Mar	0700	37.2	0.00
	0800	40.5	0.00
	0900	43.5	0.00
	1000	45.3	0.00
	1100	47.5	0.00
	1200	48.7	0.00
	1300	49.5	0.00
	1400	49.8	0.00
	1500	49.6	0.00
	1600	48.9	0.00
	1700	48.2	0.00
17 Mar	0700	40.6	0.00
	0800	41.0	0.00
	0900	41.7	0.00
	1000	43.0	0.00
	1100	44.2	0.00
	1200	46.0	0.00
	1300	49.8	0.00
	1400	51.3	0.00
	1500	53.1	0.00
	1600	54.0	0.00
	1700	54.0	0.00

APPENDIX C. SOIL MOISTURE

Date: 16 Mar 09			
Times: 0700 through 1800			
Probe Location	**Layer, in.**	**AM Reading, %**	**PM Reading, %**
Wet area	0 to 6	NA	NA
	6 to 12	NA	NA
	12 to 24	NA	NA
	24 to 36	NA	NA
	36 to 48	NA	NA
Wooded area	0 to 6	NA	NA
	6 to 12	NA	NA
	12 to 24	NA	NA
	24 to 36	NA	NA
	36 to 48	NA	NA
Open area	0 to 6	NA	5.8
	6 to 12	NA	11.8
	12 to 24	NA	14.6
	24 to 36	NA	23.8
	36 to 48	NA	25.1
Calibration lanes	0 to 6	9.4	9.3
	6 to 12	16.1	16.2
	12 to 24	20.2	20.0
	24 to 36	28.1	28.5
	36 to 48	36.2	36.7
Blind grid/moguls	0 to 6	NA	NA
	6 to 12	NA	NA
	12 to 24	NA	NA
	24 to 36	NA	NA
	36 to 48	NA	NA

Date: 17 Mar 09			
Times: 0700 through 1800			
Probe Location	**Layer, in.**	**AM Reading, %**	**PM Reading, %**
Wet area	0 to 6	NA	NA
	6 to 12	NA	NA
	12 to 24	NA	NA
	24 to 36	NA	NA
	36 to 48	NA	NA
Wooded area	0 to 6	NA	NA
	6 to 12	NA	NA
	12 to 24	NA	NA
	24 to 36	NA	NA
	36 to 48	NA	NA
Open area	0 to 6	6.6	6.5
	6 to 12	12.7	12.8
	12 to 24	14.9	15.1
	24 to 36	24.2	24.1
	36 to 48	25.5	25.4
Calibration lanes	0 to 6	NA	NA
	6 to 12	NA	NA
	12 to 24	NA	NA
	24 to 36	NA	NA
	36 to 48	NA	NA
Blind grid/moguls	0 to 6	NA	NA
	6 to 12	NA	NA
	12 to 24	NA	NA
	24 to 36	NA	NA
	36 to 48	NA	NA

APPENDIX D. DAILY ACTIVITY LOGS

Date, 09	No. of People	Area Tested	Status Start Time	Status Stop Time	Duration, min.	Operational Status	Operation Status - Comment	Track Method	Pattern	Field Conditions	
16 Mar	6	CALIBRATION LANES	1220	1330	70	INITIAL SETUP	INITIAL MOBILIZATION	GPS	LINEAR	MUDDY	CLOUDY
16 Mar	6	CALIBRATION LANES	1330	1350	20	CALIBRATION	CALIBRATION	GPS	LINEAR	MUDDY	CLOUDY
16 Mar	6	CALIBRATION LANES	1350	1500	70	COLLECTING DATA	COLLECTING DATA 1.5-METER LINE SPACING	GPS	LINEAR	MUDDY	CLOUDY
16 Mar	6	CALIBRATION LANES	1500	1510	10	DOWNTIME DUE TO EQUIPMENT MAINTENANCE CHECK	DOWNLOAD DATA	GPS	LINEAR	MUDDY	CLOUDY
16 Mar	6	OPEN FIELD	1510	1640	90	COLLECTING DATA	COLLECTING DATA 1.5-METER LINE SPACING, INDIRECT FIRE	GPS	LINEAR	MUDDY	CLOUDY
16 Mar	6	OPEN FIELD	1640	1650	10	CALIBRATION	CALIBRATION	GPS	LINEAR	MUDDY	CLOUDY
16 Mar	6	OPEN FIELD	1650	1705	15	DAILY START, STOP	EQUIPMENT BREAKDOWN	GPS	LINEAR	MUDDY	CLOUDY
17 Mar	6	OPEN FIELD	0740	0805	25	DAILY START, STOP	SET UP EQUIPMENT	GPS	LINEAR	MUDDY	CLOUDY
17 Mar	6	OPEN FIELD	0805	0915	70	COLLECTING DATA	COLLECTING DATA 1.5-METER LINE SPACING, INDIRECT FIRE	GPS	LINEAR	MUDDY	CLOUDY
17 Mar	6	OPEN FIELD	0915	0925	10	DOWNTIME DUE TO EQUIPMENT MAINTENANCE CHECK	DOWNLOAD DATA	GPS	LINEAR	MUDDY	CLOUDY
17 Mar	6	OPEN FIELD	0925	1100	95	COLLECTING DATA	COLLECTING DATA 1.5-METER LINE SPACING, INDIRECT FIRE	GPS	LINEAR	MUDDY	CLOUDY
17 Mar	6	OPEN FIELD	1100	1105	5	DOWNTIME DUE TO EQUIPMENT MAINTENANCE CHECK	CHANGE BATTERY	GPS	LINEAR	MUDDY	CLOUDY
17 Mar	6	OPEN FIELD	1105	1200	55	COLLECTING DATA	COLLECTING DATA 1.5-METER LINE SPACING, INDIRECT FIRE	GPS	LINEAR	MUDDY	CLOUDY
17 Mar	6	OPEN FIELD	1200	1215	15	DOWNTIME DUE TO EQUIPMENT FAILURE	POOR SATELLITE READINGS	GPS	LINEAR	MUDDY	CLOUDY

Date, 09	No. of People	Area Tested	Status Start Time	Status Stop Time	Duration, min.	Operational Status	Operation Status - Comment	Track Method	Pattern	Field Conditions	
17 Mar	6	OPEN FIELD	1215	1350	95	COLLECTING DATA	COLLECTING DATA 1.5-METER LINE SPACING, INDIRECT FIRE	GPS	LINEAR	MUDDY	CLOUDY
17 Mar	6	OPEN FIELD	1350	1355	5	DOWNTIME DUE TO EQUIPMENT MAINTENANCE CHECK	CHANGE BATTERY	GPS	LINEAR	MUDDY	CLOUDY
17 Mar	6	OPEN FIELD	1355	1535	100	COLLECTING DATA	COLLECTING DATA 1.5-METER LINE SPACING, INDIRECT FIRE	GPS	LINEAR	MUDDY	SUNNY
17 Mar	6	OPEN FIELD	1535	1545	10	DOWNTIME DUE TO EQUIPMENT MAINTENANCE CHECK	DOWNLOAD DATA	GPS	LINEAR	MUDDY	SUNNY
17 Mar	6	OPEN FIELD	1545	1620	35	COLLECTING DATA	COLLECTING DATA 1.5-METER LINE SPACING, INDIRECT FIRE	GPS	LINEAR	MUDDY	SUNNY
17 Mar	6	OPEN FIELD	1620	1625	5	CALIBRATION	CALIBRATION	GPS	LINEAR	MUDDY	SUNNY
17 Mar	6	OPEN FIELD	1625	1730	65	DEMOBILIZATION	DEMOBILIZATION	GPS	LINEAR	MUDDY	SUNNY

APPENDIX E.   REFERENCES

1. Standardized UXO Technology Demonstration Site Handbook, DTC Project No. 8-CO-160-000-473, Report No. ATC-8349, March 2002.

2. Aberdeen Proving Ground Soil Survey Report, October 1998.

3. Data Summary, UXO Standardized Test Site: APG Soils Description, May 2002.

4. Yuma Proving Ground Soil Survey Report, May 2003.

APPENDIX F. ABBREVIATIONS

ADST	=	Aberdeen Data Services Team
APG	=	Aberdeen Proving Ground
ATC	=	U.S. Army Aberdeen Test Center
ATSS	=	Aberdeen Test Support Services
BAR	=	background alarm rate
DMM	=	discarded military munitions
EMI	=	electromagnetic interference
EQT	=	Environmental Quality Technology
ERDC	=	U.S. Army Corps of Engineers Engineering Research and Development Center
ESTCP	=	Environmental Security Technology Certification Program
GLRT	=	generalized likelihood ratio test
GPS	=	Global Positioning System
GT	=	ground truth
HDSD	=	Homeland Defense and Sustainment Division
HEAT	=	high-explosive antitank
JPG	=	Jefferson Proving Ground
MM	=	military munitions
NS	=	nonstandard munition
PDOP	=	precision dilution of precision
POC	=	point of contact
QA	=	quality assurance
QC	=	quality control
ROC	=	receiver-operating characteristic
RTK	=	real-time kinematic
S	=	standard munition
SERDP	=	Strategic Environmental Research and Development Program
USAEC	=	U.S. Army Environmental Command
UXO	=	unexploded ordnance
YPG	=	U.S. Army Yuma Proving Ground

APPENDIX G. DISTRIBUTION LIST

DTC Project No. 8-CO-160-UXO-021

Addressee	No. of Copies
Commander U.S. Army Aberdeen Test Center ATTN: TEDT-AT-SLE (Mr. J. Stephen McClung) 400 Colleran Road Aberdeen Proving Ground, MD 21005-5059	1
Commander U.S. Army Environmental Command ATTN: IMAE-RTA (Ms. Kimberly Watts) 314 Longs Corner Road Aberdeen Proving Ground, MD 21010-5401	1
Environmental Security Technology Certification Program (ESTCP) ATTN: Mr. Herb Nelson 901 North Stuart Street, Suite 303 Arlington, VA 22203	1
Defense Technical Information Center 8725 John J. Kingman Road, Suite 0944 Fort Belvoir, VA 22060-6218	PDF

Secondary distribution is controlled by Commander, U.S. Army Environmental Command, ATTN: IMAE-RTA.